FDsys Update

October 19, 2010

FDsys

FDsys Background

- FDsys automates the collection and dissemination of electronic information from all three branches of government.
- Information is:
  - Submitted directly into FDsys
  - Permanently available in electronic format
  - Authenticated and versioned
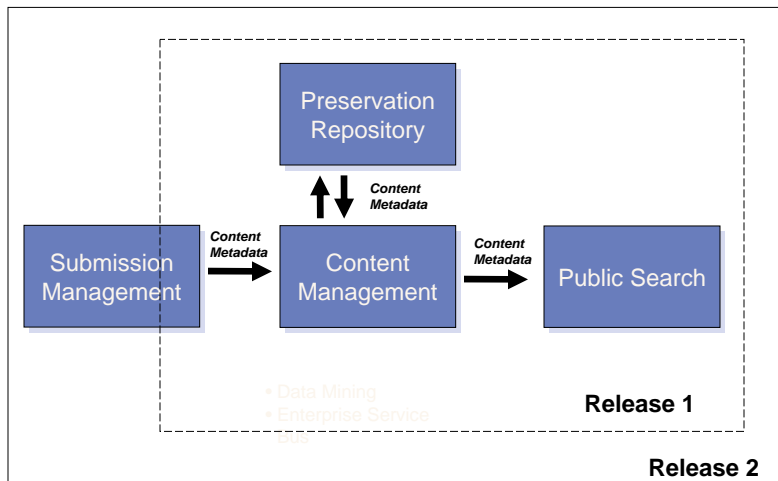 - Publicly accessible for searching and downloading

1

What is FDsys?

- FDsys is a *Content Management System*
  - FDsys securely controls digital content throughout its lifecycle to ensure content integrity and authenticity
- FDsys is a *Preservation Repository*
  - FDsys follows archival system standards to ensure long-term preservation and access of digital content
- FDsys is an *Advanced Search Engine*
  - FDsys combines extensive metadata creation with modern search technology to ensure the highest quality search experience

2

FDsys Core System Components

Preservation Repository

Content Metadata

Content Metadata

Submission Management

Content Management

Content Metadata

Public Search

- Data Mining
- Enterprise Service

Release 1

Release 2

3

Release 1 Goals

- Establish the foundational infrastructure
- Establish preservation repository
- Replace current public site
- Perform large scale data migration
- Provide operational continuity for the system

4

Release 1 – Current Status

- Foundational infrastructure in place
  - FDsys search component live in January 2009
  - FDsys CMS component live in March 2009
    - Provides a place where text, locator, PDF, etc can be stored in one location
- Preservation repository in place
  - Preservation repository live with CMS launch in March 2009
  - World class preservation repository built on OAIS model
  - Repository provides for sufficient control to ensure long-term preservation and access

5

Release 1 – Current Status (cont'd)

- FDsys public interface in place
  - Public site live with search component launch in January 2009
  - Worked closely with stakeholders on interface, including the library community, agency customers, and congressional customers
    - Customizations made based on collection
    - Congressional committees designed browse pages
- System Continuity of Access (COA) site is operational
  - COA failover tests successfully completed and Authority to Operate (ATO) has been granted
  - With COA in place, GPO can maintain public access to historic and new content in the event of an emergency
- Full system failover (all FDsys components) planned to be operational in December 2010

6

Release 1 – Current Status (cont'd)

- Data migration process is completed
  - All 40 content collections completed in June 2010
  - All non-Wais content migration completed in September 2010
    - ~100 individual publications identified on GPO web servers
  - Introduced significant improvements during migration
    - Automated metadata extraction process for content collections developed to support search and preservation
    - Data inconsistencies handled by parser and metadata editor
    - Digital signatures added to PDF files in content collections
- Remaining quarantined packages are being worked
  - Extensive analysis of multiple data storage servers identified the full scope of GPO's data universe
  - Review uncovered issues with a subset of content during migration
  - Less than one percent of content packages remain to be submitted to FDsys during transition process

7

Additional FDsys Projects

- Released a new publication, *Daily Compilation of Presidential Documents* in February 2009
- In support of the Office of the Federal Register, developed XML content and made available through Data.gov
  - FR XML released in September 2009
  - CFR XML released December 2009
- Worked closely with the Office of the Law Revision Counsel of the U.S. House of Representatives to bring the yearly virtual main editions of the United States Code collection into FDsys in XHTML and PDF formats
- Developed a pilot project for the digitized Statutes at Large, demonstrating converted content preservation and access in FDsys

8

FDsys and Open Government

- FDsys is based on the concepts of interoperability and reuse
- FDsys is making a difference
  - FDsys content available in all major search engines
  - Citizens are finding new ways to interact with key public content through XML data
    - Includes sites like FedThread.org and govpulse
  - Other government sites rely on FDsys to enrich their user experience
    - FR 2.0 (federalregister.gov) developed by OFR and GPO as an XML-based edition of the daily Federal Register
    - Regulations.gov uses FDsys metadata to power their public search, eliminating the need for OFR support
    - Science.gov is able to integrate GPO collections into the federated search gateway to government science information

9

Completing Release 1

- GPO Access to FDsys transition is in process
  - Key activities have been identified and a completion plan has been developed
    - GPO Access redirects
    - Getdoc link redirects
    - Quarantined package completion
    - Technical planning for GPO Access shutdown
- Once completed, FDsys will assume role as GPO's electronic system of record
  - Sunset planned for December 2010

10

Release 2 Goals

- Release 2 is underway
  - Key new component is external content submission
    - Currently working with Congressional Publishing Services on Congressional Content Workflows
  - Beyond submission, additional functionality will extend access functionality and data usability
    - With the foundational system in place, a strong emphasis will be on bringing more content into FDsys
  - Release 2 will be implemented over four planned major releases
    - Will include new features as well as identified system fixes
    - Minor releases may occur between major releases if needed
    - New features will continue to be added to the FDsys roadmap as needs are identified by stakeholders
      - Priorities may change based on identified needs
      - Proposed changes to planned roadmap will be communicated with stakeholders

11

R2 Priorities

- Congressional content submission
- Collection updates
    - Public Papers of the President in XML
    - New US Government Manual
    - House Journal historical and new content
    - Report of the Secretary of the Senate collection
- Harvested content support
    - Harvested Content Strategic Roadmap
    - Collection development- Coastal Zone Information Center collection
- ILS integration
- Introduction of preservation processes
    - TRAC audit
    - Format Validation (use of JHOVE)
- Technology improvements to FDsys
    - Replacement of end of life technologies
    - Implementation of Enterprise Service Bus to support ILS integration

12

Features for R2 Group A Release

Feature Name	Feature Description
Additional Renditions for Public Papers of the President	Support for the submission, processing, and providing public access to new renditions for the Public Papers of the President in XML
AOUSC Pilot (awaiting JCP approval)	Support for the submission, processing, and providing public access to content from the Administrative Office of the US Courts
Congressional Committee Browse Pages Enhancements	Enhancements to existing committee browse pages
Enhanced Metadata Editor (XME)	Completion of all remaining requirements and fixes for the FDsys metadata editor (XME)
House Journal Collection Update	Adding legacy content and new content to the existing House Journal collection
System Continuity of Operations (COOP)	Includes requirements needed to stand up a COOP instance of FDsys
WebLogic Upgrade	Oracle Application Server is at end of life. Need to upgrade application servers for FDsys

13

Related Projects

- Composition System Replacement
  - GPO moving to a more flexible, interoperable framework for composing and preparing essential documents of the Federal Government
  - Will leverage XML to accept and compose files in standard formats and provide final XML inputs to FDsys, allowing for more powerful search and promote content reuse.
  - Design document to be completed by the end of the year
- Agency ordering
  - Will modernize ordering processes for agency and congressional customers
  - Parallel project to FDsys content submission

14